DOCUMENT RESUME

ED 152 833 TM 007 042

AUTHOR Backman, Margaret E.
TITLE The Effectiveness of Relevant Item Content in Reducing Test Bias.
PUB DATE 74
NOTE 13p.; Paper presented at the Annual Conference of the National Rehabilitation Association (Las Vegas, Nevada, October, 1974) ; Best Copy Available

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Achievement Tests; *Adult Basic Education; Adults; *Basic Skills; Culturally Disadvantaged; *Educationally Disadvantaged; *Item Analysis; Occupational Guidance; Occupational Tests; Practical Mathematics; Racial Differences; Reading Ability; Tape Recordings; *Test Bias; Test Construction; Test Items; Verbal Tests
IDENTIFIERS *Fundamental Achievement Series

ABSTRACT

For decades there have been attempts to develop tests that are fair to people from different backgrounds. The Fundamental Achievement Series (FAS) is a recent attempt to develop a test that is fair to those considered disadvantaged. In addition to the traditional kinds of items, the FAS incorporates relevant items, tapping verbal and numerical abilities required in common everyday situations. A comparison of responses to relevant items and traditional ones suggests that, when testing disadvantaged adults, it is important to use the more relevant items. Item analysis indicated that on the FAS numerical test, the relevant items were easier, yet had higher discrimination indices. On the verbal test, the relevant items worked as well as the traditional items. The FAS is not free of cultural bias; differences in test performance do occur between blacks and whites. Performance on both parts of the FAS is related to reading ability. However, adults with poor reading skills have an advantage on the FAS that is not present on traditional paper and pencil tests: the instructions for the FAS are played on a tape recorder. (Author/CTM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED152833

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Margaret E. Backman

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

The Effectiveness of Relevant Item Content in Reducing Test Bias

Margaret E. Backman, Ph.D.
Director of Vocational and Social Science Research
ICD Rehabilitation and Research Center
340 East 24th Street
New York, New York 10010

BEST COPY AVAILABLE

TM007 042

Based on a paper presented at the annual conference of the National Rehabilitation Association, Las Vegas, Nevada, October 1974.

ABSTRACT

The Effectiveness of Relevant Item Content

in Reducing Test Bias

For decades there have been attempts to develop tests that are fair to people from differing backgrounds. The Fundamental Achievement Series (FAS) is a recent attempt to develop a test that is fair to those considered disadvantaged. In addition to the traditional kinds of items, the FAS incorporates relevant items, tapping verbal and numerical abilities required in common everyday situations. A comparison of responses to relevant items and traditional ones suggests that, when testing disadvantaged persons, it is important to use the more relevant kinds of items. Item analysis indicated that on the FAS Numerical test the more relevant items worked psychometrically better than the traditional ones, i.e., were easier yet had higher discrimination indices. On the Verbal test, the relevant items worked as well as the traditional items. It was noted, however, that the FAS is not free of so-called cultural bias: differences in test performance do occur between Blacks and Whites; also, performance on both parts of the FAS is related to reading ability.

The Effectiveness of Relevant Item Content

in Reducing Test Bias

There have been numerous attempts in the past to develop tests that would be considered "culture-free." For a number of reasons such attempts have not been successful: Tests are in effect samples of behavior in society; thus, past experience does influence present functioning on tests. To eliminate items that are related to experience would leave you with a test that would be measuring unimportant, trivial behavior.

A more recent trend has been to develop tests that include items considered fair to all groups being tested, (e.g., the advantaged and disadvantaged, males and females). To a large extent the content of these tests is more relevant to the everyday experiences of the persons tested. The content of the items depends upon cultural influences, but as much as possible, the items are supposed to be equally familiar to all groups concerned. This means that either the groups perform equally well on all items, or that an equal number of items favoring specific groups appear in the test. These tests do not attempt to eliminate cultural influences, but attempt to sample cultural experiences that are common to those being tested.

Such tests, if successful, would be extremely useful in vocational and educational guidance. It is important, however, that these new tests be examined carefully to see if they indeed function in the way the test publishers say they do. The purpose of the present study was to compare the effectiveness of so-called relevant items to items traditionally found in paper-and-pencil tests.

The Fundamental Achievement Series.

The specific test of interest in this paper is the Fundamental Achievement Series (FAS) published by the Psychological Corporation (1968), and reviewed by Lewis R. Aiken in the Journal of Educational Measurement (1970).

The FAS contains a Verbal and Numerical section, each of which can be administered in 30 minutes. A tape recorder is used for presenting the questions and directions, and for controlling the timing; the tape has the advantage of eliminating the necessity for an examinee to read instructions.

The FAS is designed to cover a range of ability from basic literacy to somewhat above eighth grade. The manual states that the tests are intended for use in the employment or educational placement of adults and adolescents who may not have had the usual exposure to formal education.

The Verbal test includes items that measure the ability to read signs, to add prices that appear on a restaurant menu, to find names on apartment house lists, to find numbers in a telephone book, to copy sentences in long hand, and to understand and remember details from brief oral announcements. More than half of the test, however, includes the more traditional kinds of items found in paper-and-pencil tests: recognizing misspelled words, selecting the synonyms for given words, and picture vocabulary.

The Numerical test includes items that measure the ability to read and write numbers, to tell time, to add the amount of coins, to know simple facts, such as the number of inches in a foot, to measure with a ruler, to calculate the area of a rectangle, to add a sales slip, to write a check, to read a calendar, and to read charts. In addition, thirty of the sixty-nine items in

the Numerical test are similar to those found in traditional paper-and-pencil tests: arithmetic computation, and reading problems involving arithmetic reasoning.

The authors of the test feel that the inclusion of items related to everyday activities should increase the motivation of testees. In addition, although overall test scores are more reliable than performance on individual items, for diagnostic information attention might be given to certain parts of the test. For example, certain jobs might require an applicant to take messages and to write legibly, or to add up a restaurant check. Such diagnostic use of the test, however, is limited, since a small number of items represents only a small sample of a person's behavior, and may not be a reliable indication of a person's ability.

Using the FAS in a rehabilitation facility: A study.

There is still some question as to whether or not the FAS has any advantage over other traditional tests used with disadvantaged persons. To examine this question the FAS was given to a sample of clients (n = 61) in a comprehensive rehabilitation center in New York City. All clients tested were at the center for the purpose of vocational evaluation. A little more than half of the clients (59%) were women; 44% were Black, and 30% were Spanish-speaking. The ages ranged from less than 20 to 56 years, with more than half of the clients (57%) between 20 to 34 years of age. Over half of the clients (61%) had between 10 to 12 years of education.

It was not clear from the testing to what extent the FAS was able to increase motivation and reduce frustration. To a large extent the FAS was still seen as a paper-and-pencil test, and had some of the limitations of such tests.

Many clients had difficulty marking the responses in the appropriate spaces. Although testees are to write in the booklets and not on answer sheets, some clients still had difficulty marking correctly. For example, some put X's on the pictures in the picture vocabularly section rather than in the box underneath the picture. These kinds of errors were common, reflecting a lack of experience in test taking, and perhaps a difficulty in understanding and following instructions.

Another problem area is the tape recording itself. In a few instances the actual pronunciation is not clear (e.g., "Cliff Street" on the bus sign). Also the person's voice has a definite regional accent, which was commented upon by the New York City clients. It is not known what effect an unfamiliar accent might have on attention and comprehension.

Also, despite the use of the tape recorder for giving instructions, both the Verbal and Numerical parts of the FAS rely to a large extent on reading ability. The manual reports a correlation of .94 (n = 43) between the FAS Verbal and the Gates-MacGinitie Reading Test. Correlations between the FAS Verbal test and the Verbal Reasoning test of the Differential Aptitude Tests (DAT) for eighth graders ranged from .43 - .57. Correlations between the FAS Numerical test and the Verbal Reasoning test of the DAT ranged from .22 - .64. The high correlations with traditional tests may

in part reflect the large number of traditional items included in the FAS.

Comparison of relevant and traditional items: Verbal

The question then arises: How successful are the more relevant items on the FAS, as compared to the more traditional items? Table 1 may serve to answer this question.

Insert Table 1 about here

Table 1 reports the median level of difficulty for sets of items measuring certain skills and also the median discrimination index (point-biserial r) for each set of items. The range of the difficulty and discrimination indices for the sets of items is also given.

Thus, on the first six items of the Verbal test 80% of the clients, on the average, were able to read signs, such as "Smoking Prohibited." The actual percentages of persons getting the items correct ranged from 67% - 87% over the six items. The average point-biserial r, using the total score on the Verbal test as criterion, was .54. These results indicate that for these six items, persons having the highest total scores on the FAS Verbal test also tended to be the ones to get these six items correct. Thus, these particular items seem to be doing their job, that is, identifying persons with good verbal ability.

The overall conclusion for examining Table 1 is that the relevant items (items 1-34) on the Verbal test seem to be functioning as well as the more traditional items (items 35-100). Items 25-34 seemed to be quite difficult, however, and do not discriminate as well as the rest of the items in the test. In this part of the test announcements are made by the voice on the tape, and include details, such as the place and time of a social event. After each announcement, the examinees are asked to recall specific facts. The questions about the announcements require attention to detail, concentration, verbal comprehension, and short-term memory. Performance could easily have been affected by anxiety level.

The traditional items where one is to select the synonym of a word were also somewhat difficult, but did work well as far as identifying the more verbally proficient clients.

In summary, on the Verbal test the inclusion of more relevant material may not contribute any more psychometric information than conventional items; on the other hand, the inclusion of relevant items did not adversely affect test performance. It is possible that motivational level may have been influenced by the use of the more relevant content, but such content has not been shown to have had a differential effect on performance on the Verbal test.

Comparison of relevant and traditional items: Numerical

For the Numerical test another picture seems to emerge. The easiest items and the items with the highest discrimination indices were those that were more relevant and whose content was more familiar to the examinees

(e.g., telling time, adding coins, recognizing numbers, and general information such as the number of inches in a foot). Reading a mileage chart seemed to be very difficult, although the clients were able to read a postal chart and a sales chart. However, the items involving chart reading in general did not discriminate well between those with high and low total scores on the Numerical test.

The traditional reading problems (arithmetic reasoning) were very difficult, with only 12% getting an item correct on the average. These problems reflect both reading ability, as well as level of arithmetic reasoning. Performance on such tests is very much influenced by educational experience and training.

Although more research needs to be done in this area, it appears that for disadvantaged adults it is important to include items on tests of numerical ability that are directly related to everyday experience. Such items seem to be better for identifying persons who can work in areas requiring some ability to work with numbers. Such items may also be useful for selecting persons who could benefit by additional schooling or remedial instruction.

Summary: Using the FAS in vocational and educational guidance.

The FAS has been published as a research edition and rightly so. More data must be collected on this test before its value in the field can adequately be judged. A revised form will undoubtedly be published at a later date.

The FAS is a good attempt to construct an achievement test suitable for disadvantaged adults. The FAS has the advantage that it is easy to administer and takes only an hour of the client's time. Thus, it is a relatively quick and efficient way of getting an estimate of a client's basic verbal and numerical ability; a combined score can give an indication of the person's general level of achievement. It should be noted that in vocational counseling a verbal and numerical score are not necessarily sufficient to predict success in one job as compared to another. Such scores do, however, give an indication of a person's present level of ability and suggest possible levels of entry for jobs or further opportunities for educational training.

Another advantage of the FAS is that instructions have been recorded on a tape. Thus, persons with poor reading skills should have an advantage on this test over traditional paper-and-pencil tests. Reading ability does play a large part, however, in being able to answer questions on both the Verbal and Numerical tests.

What makes the FAS worth further investigation despite the limitations discussed previously, is the use of items based on everyday experience. Such items should make the test more interesting to adults who have very practical concerns and want to see the relevancy of tests they are taking. In addition, the results of the study reported here suggest that the more relevant everyday items may be very important in tests of numerical ability, particularly when testing disadvantaged persons. It should be noted, however, that although the test was designed to contain material familiar to both the advantaged and disadvantaged, results of studies reported in the FAS manual

reveal that differences in test performance between Blacks and Whites have not been eliminated.

Although there are limitations to the FAS, it is probably one of the best around for its stated purposes. Individual users should, however, conduct their own studies to see how useful the test is for their specific needs.

References

Aiken, L. R. Fundamental achievement series. Journal of Educational Measurement, 1970, 7(2), 134-136.

Bennett, G. K., & Doppelt, J. D. Fundamental achievement series. New York: The Psychological Corporation, 1968.

Table 4

Performance of Educationally Disadvantaged Adults on Relevant and Traditional Items for the Fundamental Achievement Series (Form B)

Items	Options	Description of Content	Difficulty[a]		Discrimination[b]	
			Median	Range	Median	Range
		Verbal test[c]				
Relevant						
1-6	4	Public signs	80	67 - 87	.54	.46 - .56
7-8	4	Bus signs	70	62 - 77	.34	.33 - .35
9-13	open	Menu	57	13 - 74	.55	.20 - .56
14-18	12	Apartment house list	90	79 - 93	.50	.41 - 67
19-21	9	Telephone book	80	80 - 92	.54	.46 - .54
22-24	open	Copy sentences	62	61 - 70	.57	.56 - .57
25-34	open	Announcements	30	16 - 48	.38	.29 - .42
Traditional						
35-58	2	Spelling	70	32 - 95	.50	.02 - .69
59-76	4	Picture vacabulary	81	59 - 98	.41	.00 - .53
77-100	4	Synonymns	59	36 - 85	.57	.31 - .70
		Numerical test[d]				
Relevant						
1	4	Number recognition	93	93	.46	.46
2-3	4	Time (clock) recognition	89	89	.58	.57 - .59
4-9	open	General information	61	16 - 80	.50	.44 - .57
10-13	open	Adding coin values	89	87 - 93	.46	.43 - .52
14-22	4	General information	75	28 - 98	.39	-.12 - .54
23	4	Reading a ruler	33	33	.08	.08
24	4	Computing area	11	11	.19	.19
25-26	open	Adding a restaurant bill	55	54 - 56	-.12	-.19 - (-).05
27-28	open	Writing a personal check	36	34 - 38	.12	.06 - .17
29-31	open	Reading a calendar	46	36 - 56	-.09	-.10 - (-).08
32-33	open	Reading a chart (graph)	50	43 - 56	-.07	-.09 - .05
34-36	open	Reading a postal chart	48	26 - 49	.03	.03 - .15
37-38	open	Reading a mileage chart	02	00 - 04	-.12	-.24 - .00
39	open	Recognizing "radius"	10	10	.22	.22
Traditional						
40-59	open	Numerical computation	33	03 - 56	.05	.15 - .29
60-69	open	Word problems	12	02 - 41	.14	.00 - .25

[a] Percent correct

[b] Point-biserial r

[c] N = 61, $\bar{x}$ = 66.62, SD = 19.41

[d] N = 61, $\bar{x}$ = 40.59, SD = 13.29